Alice in "Bio-land": engineering challenges in the world of Life-Sciences

Alice in "Bio-land": engineering challenges in the world of Life-Sciences

Alfredo Benso[1], Enrico Bucci[2], Stefano Di Carlo[1], Gianfranco Politano[1], and Alessandro Savino[1]

[1]Department of Control and Computer Engineering of Politecnico di Torino, Corso Duca degli Abruzzi 24, Torino, Italy - e-mail: {firstname.lastname}@polito.it

[2]BioDigitalValley S.r.l., Pont St. Martin, Aosta, Italy - e-mail: enrico.bucci@biodigitalvalley.com

***Abstract*—Alice is an engineer who ventures into the research world of Life Sciences. To her eyes, Life Sciences researchers work backwards compared to what happens in her world. It appears that their research methodology has a number of issues that may limit its potential. Nevertheless, she also becomes aware that a different set of problems arises if her own traditional top-down engineering approach is applied to Life Sciences. This paper discusses how we see the role of Systems and Computational Biology as a fundamental methodological "middle-ground" between these two (apparently) distant worlds.**

I. Introduction

The typical way the Life Science world sees the role of Computer Science is a "tool" to analyze laboratory data. At a first glance this role is a reasonable one, since computers play a significant part in interpreting and statistically analyzing biological data. Computers started to become necessary (and not only sufficient) when Biologists were not able anymore to manually "infer" plausible hypotheses and models from the huge amount of data becoming available thanks to high-throughput biotechnologies, and even more from the use of the web as a main "big data" repository. Nevertheless, there is a wide gap between what computer science is doing for the Life Sciences research community, and what it could do. This is due not to technical reasons, but to a methodological problem based on a bottom-up approach where where computation is mostly used to "make sense" of lab data. In this paper we ask ourselves: is this the only plausible, or the most efficient and effective, approach for the engineeering and computer science world to assist the type of work Life Science researchers are doing? What is it exactly, that they are doing? To some extent, in our opinion, the effort to formalize the results of an experiment should be comparable to the one required for the data analysis done to design the experiment itself; nevertheless this habit requires an engineering frame of mind that is often missing or neglected in the Life Science community.

II. Biology is Reverse Engineering

First of all, let's try to better understand and formalize the type of research most scientists are doing in the Life Sciences world. Biologists are not designing a system like most engineers do. They are instead trying to understand how a certain biological system works. In the engineering world, this task is known as Reverse Engineering, a concept very clear to engineers but not so easy to understand for other scientists, especially in its implications. "*Reverse engineering is the process of discovering the functional principles of a device, object, or system through analysis of its structure, function, and operation*" [6]. Life Sciences are "the" Reverse Engineering sciences by definition: their task is to understand, model, and predict the dynamics of the biological systems that allow, define, and regulate life.

One key concept to understand about reverse engineering is that *designing a system* and *reverse engineering* it are two opposite tasks whose complexity may differ in orders of magnitude. Considering this, the question we now ask is: should the complexity (in mathematical terms) of the system under investigation drive the methodological approach followed to reverse engineer it?

Answering this question is fundamental because it could have important consequences not only on the way research is planned, but already on how IT Professional curricula that focus on Life Sciences are designed and organized. The choice of a methodological approach is not only a consequence of the notions scientists learned during their studies, but even more of the *forma mentis* that a particular course of studies shaped in each of them.

Since we started applying Computer Engineering to Life Sciences, we observed a definite cultural gap between biologists and engineers. Biologists seem to be very comfortable with ambiguity, whereas engineers always expect to achieve established knowledge and certainty. The methodological differences between these two are planted during undergraduate and graduate studies. Engineers are trained to work top-down, with an approach which is focused towards modeling as much as possible, stopping only when the model is detailed enough to make all the available results and knowledge smoothly work together. Biologists are taught a bottom-up approach, analytically and precisely studying very specific and controlled problems, but lacking reliable methodological tools that allow them the generalization of their findings and the understanding of the higher-level dynamics of the systems under study. What is important to understand is if the success of one approach or the other is an "absolute" characteristic of the approach itself, or if it depends on the complexity and properties of the system under study.

A. An Example: Nyquist, a movie plot, and genetic regulation

To better understand why the question posed in the previous section is so important, let's make a very simple example. In

most engineering fields, a common problem is the determination of the sampling frequency necessary to reconstruct a signal (of any kind). The Nyquist-Shannon law comes to help basically stating that, to obtain a reliable signal reconstruction, the sampling frequency has to be greater than the double of the signal frequency (or signal dynamic). Let's "translate" this into a simpler example, maybe not technically completely correct, but easier to understand for those readers that are not familiar with signal reconstruction problems.

Let's assume that our task is to figure out the plot of a movie, without seeing the whole movie, but only screenshots of it. The question is: how often should we grab a screenshot (the *sampling rate*) to be able to reliably reconstruct the plot of the movie? The answer only depends on how often a significant event takes place in the plot. If, for example, we assume that two consecutive significant events never take place less than 10 minutes apart, then the Nyquist law tells us that we need to get a screenshot at most every five minutes in order to be able to reliably infer anything about the plot.

We used this example, despite not being completely theoretically correct, because it is closely related to a typical "bioinformatic" problem, that is the understanding of the wide range of mechanisms that are used by cells to increase or decrease the production of specific gene products (proteins). The "expression" of a gene is the process by which the information encoded in a gene is used to synthesize a functional gene product, e.g. a protein. Genes are not expressed all the time; instead, the level of expression of a gene depends on the expression level and the interactions among the products of several other genes. The most effective way to represent this network of interactions is by using graphs, where each node is a gene, and each edge connecting two genes represents a direct regulatory interaction between them. The topologies of these networks are static (they do not change in time), but the expression profile of each gene does change in time depending on the status of the other nodes, and represets the network dynamics. Biotechnologies (e.g., microarrays) allow to take "snapshots" of the expression levels of thousands of genes at different sampling rates. The problem can be then formalized as: given a set of snapshots of the expression levels of a group of selected genes taken at variable time intervals (usually minutes or hours), is it possible to reconstruct the network of interactions that models the system dynamics? Using the analogy introduced before, each gene expression profile is like a movie screenshot, and the final network topology (and consequently, dynamics) is the movie plot.

Before even considering the different approaches proposed to "build" the gene interaction network from the expression profiles, we ask the following question: if the current biotechnologies allow to photograph the gene expression for example every 10 minutes, will this sampling frequency be compatible, following the Nyquist-Shannon law, with the network operating frequency? Because, in this example, the network would be "reconstructable" only if its "frequency" is in the range of tens of minutes. So, what is the frequency at which a gene regulatory network works? In our discussion with several researchers we still haven't found a clear and definite answer. No one, to the best of our knowledge, ever

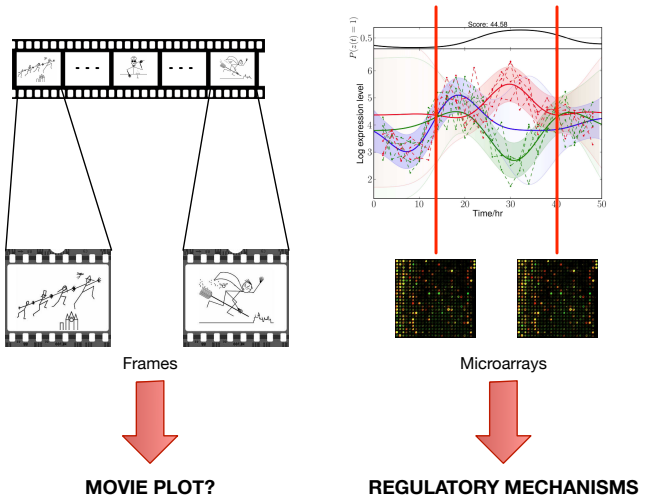

Figure 1. Sampling frequencies

seriously considered this question in these terms. Apparently the dynamics of some genes are in the order of seconds, others work in the order of minutes, others of hours... so what is the accuracy in reconstructing a network in this way, without even taking into account the fact that if the sampling frequency is incompatible with the system dynamics, then no expression sample can be reliably used? Does it make sense to identify drug targets, associate pathologies with genes, or design a therapy with data that give no guarantee of revealing anything about the real systems' functionality? For further discussions about these issues, a recent and interesting article specifically focused on the reverse engineering of regulatory systems can be found in [1].

III. Methodologies: Bottom-up vs Top-down

The final goal of both bottom-up and top-down approaches to Reverse Engineering is to construct a reliable model of the system under investigation. Modeling is a way to encapsulate part of the real world in terms of mathematical relationships. The key points of both approaches are summarized in Figure 2.

The bottom-up approach works extremely well with linear systems. Linear systems are subject to the principle of superposition that states that the net response at a given place and time caused by two or more stimuli is the sum of the responses which would have been caused by each stimulus individually. So that if input A produces response X and input B produces response Y then input (A + B) produces response (X + Y). The direct consequence of this property is that it is possible to study the dynamics of the whole system by studying, individually, the dynamics of each component. Linear systems are easy to understand also for non-mathematicians and are easy to visualize. For this reason a large part of the Life Sciences world, and particularly the medical one, still reasons in linear terms, neglecting the fact that probably only a few of the biological systems are truly linear. The main obstacle appears when it is time to merge all experimental observations to build a higher-level model (the "up" part of the methodology).

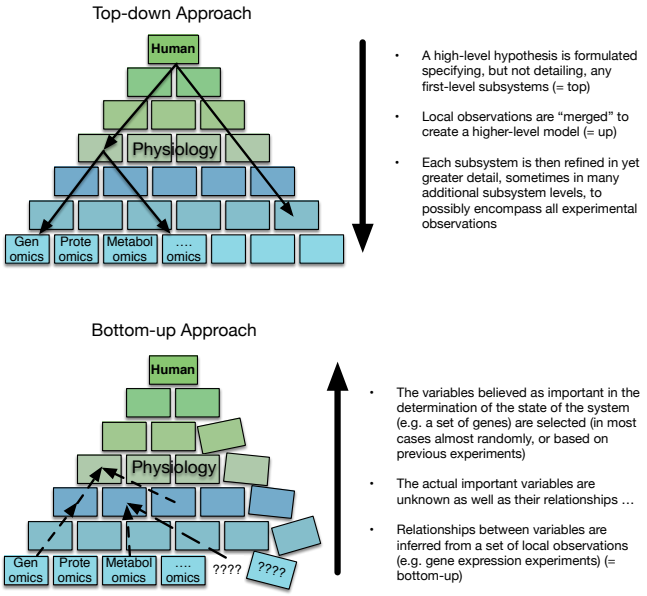

Figure 2. Top-down vs Bottom-up

If the system is not linear but complex, this step is not guaranteed to be possible since the superposition property does not hold anymore. Complex (non-linear) systems are much more difficult to understand or visualize. They consist of many diverse and autonomous (but interrelated and interdependent) components or parts linked together through many (usually dense) interconnections. They cannot be described by a single rule and, importantly, they exhibit properties that emerge from the interaction of their parts and that cannot be predicted only from the properties of each individual part. What happens is that often observations that "do not make sense" are discarded as "lab errors", or results are manipulated to make everything fit together. Nevertheless, this failure in the abstraction process may not be caused by errors made during the research, but instead by a limitation of the methodology itself, which is intrinsically unable to elucidate some of the most important functional properties of the system. In this situation, a bottom-up approach has strong limitations, clearly demonstrated by huge variations in the results of lab experiments. In the medical world, the bottom-up approach has a very misleading characteristic that makes its use very "tempting": it is able to show early insights on the system behavior. The mistake is to consider these insights reliable enough to produce a higher-level model (or even a therapy for a disease), whereas it is very likely that some key dynamics of the system are missed because of the specificity of the observations, or, worse, because the complexity itself hides the basic mechanics.

The top-down approach has its limitations too. Since it works on abstractions and inferences, the conclusions reached are often general enough to try to explain the overall mechanics but their basis often lies on computational assumptions that may be incorrect. The top-down solution risks, therefore, to provide, in the first phases, limited correlation with experimental data and, consequently, the potential benefits of the approach may not be immediately evident. There is also the risk, while refining the model, to increase its complexity to the point where it is not computationally manageable anymore. However, this risk is mitigated by the steady increase in the computational power of modern computers. The implementation cost of a top-down approach is also likely to be higher, since it may require several iterations and refinements to correctly adapt the model to the available experimental data. Moreover, often some model features may not be comprehensively confirmable by experimental evidence, and this constitutes the primary reason for the skepticism that Life Sciences researchers have towards Systems and Computational Biology. On top of this comes the fact that top-level approaches may provide models that may not be easily transposable to real applications or therapies, whereas medical and biological research need methodologies that can be quickly translated into new drugs and treatements.

IV. The role of Systems Biology

For a long time it was assumed that the two approaches discussed in the previous section could be pursued independently, and would eventually meet in the middle. This proved true in the reverse engineering of linear or almost linear systems but it showed huge limitations for complex systems. Even BigPharma companies are understanding that the integration of the two approaches is the only way to provide new momentum to the drug target discovery market that, currently, is stagnant because of the lack of new molecules and targets to test.

Systems and Computational Biology are not only the latest fashion in biology, but a necessary step to overcome the limitations of both methodological approaches and to find a "middle ground". They constitute the best option to create a feedback loop between modeling and laboratory experiments. As presented in Figure 3, Systems and Computational Biology are the disciplines that should, starting from high-level hypotheses and models, drive the experimentation and laboratory phases. By closing the methodological gap between top-down and bottom-up, they can establish a loop that has the potential of getting the best advantages of both approaches, while overcoming most of their limitations.

V. The case of Biological Networks

A very good example of how an "engineering mind" can (and did) contribute to biological research, is the field of biological networks, where graphs and network structures are used to model interactions and relationships among biological entities. These models are now more and more used to discover emergent properties among genes, proteins and other relevant biomolecules referred to specific phenotypes or diseases. Nevertheless, the idea of correlating biological properties to topological features of a network did not come directly from the Life Science world. It was actually born almost by chance thanks to the fact that two scientists, Oltavi, a cell biologist and Barabasi, a physicist, were neighbors. At the time, Barabasi had already shown that the internet is a non-random network, and that its connectivity structure influences its function. One year later, in 1999, they proved that the metabolic pathways of yeast define a network whose structure is very similar to

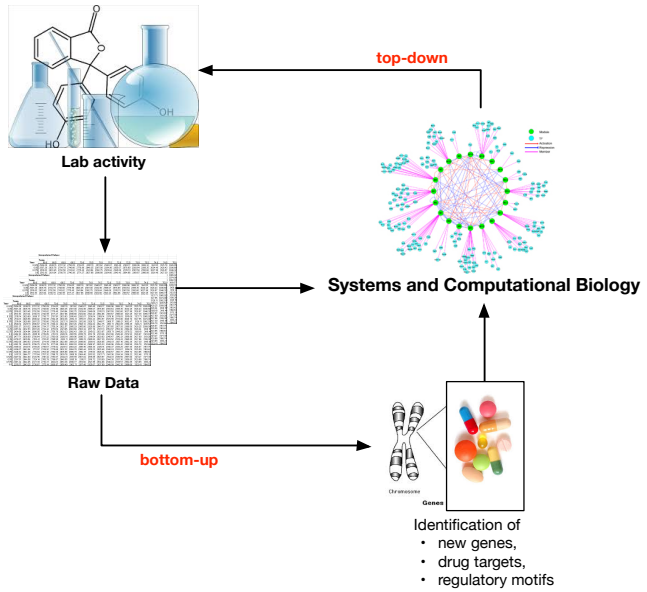

Figure 3. The role of System and Computational Biology

the internet. From that moment, several theoretical studies confirmed that biological networks share many features with other types of networks such as computer or social networks, and, even more importantly, they make several mathematical and computational methods of the graph theory applicable to biological studies [2] [11]. Probably the keystone in Network biology can be found at the turn of the century when three scientists, in a now classical Nature paper titled "Surfing the p53 network" [20], were able to explain the tumor-suppressor role of the gene p53 (which was found mutated in about 50% of cancer patients) by analyzing its network topology. They wrote that "The cell, like the Internet, appears to be a 'scale-free network'", and "one way to understand the p53 network is to compare it to the Internet". The surprising point here is that in 1999 more than 15.000 articles had already been published on p53 and its role in cancer biology, yet, after 20 years of research, some aspects of p53 biology were still missing and they could be unveiled thanks to a methodology that had nothing to do with the traditional biological research approach.

Since then, the computational analysis of biological networks has become increasingly used to mine the complexity of cellular processes and signaling pathways. From simple statistics gathered from PubMed, it appears that the scientific communities of both areas are steadily converging on this approach. The literature describing promising applications of network analysis in biology is growing at an exponential rate, as shown by the increase in the number of papers yearly indexed by Pubmed for "network biology": 407 in 2000, 1018 in 2004, 2342 in 2010, and 2558 in 2012.

To better understand the potential of Networks Biology, it is probably worth clarifying some of its basic characteristics. Since a Complex System shows properties that cannot be understood from the isolated study of its components, the main contribution of the network model is that it provides a way to study interactions.

Many types of biological networks do exist, depending on the information associated to their nodes and edges. In general, they can be classified as directed or undirected networks ([16]). In directed networks, nodes are molecules, while edges indicate causal biological interactions among nodes (e.g., transcription and translation regulations (Li et al., 2012)). Instead, in undirected networks, an edge indicates a shared property, such as sequence similarity [13], gene co-expression ([17], [3]), protein-protein interaction ([5]), or term co-occurrence in the scientific literature ([9], [8]).

In general, there are two main approaches to network analysis: *the topological analysis,* that tries to identify recurrent network sub-structures and correlate them with particular functional characteristics of the system, and the *dynamic analysis*, which focuses instead on the study of the evolution of the network in time. In the following two subsections we will show some examples of how the results of these network analysis approaches, when applied to biological networks, can actually provide biologically meaningful infromation that could not be easily identified using a straight bottom-up laboratory-based approach. One of the key points of this paper is to show how more holistic approaches have the potential not only to support the biological research, but also to boost it and to suggest new and previously unseen directions.

A. *Topological network analysis*

To give the reader a general idea of the main achievements in this field, we focus on three well known topological features of a graph: hubs, cliques, and graphlets. Their identification in biological networks allows the unveiling of interesting biological mechanisms that emerge from the *interaction* of the network components, which would be impossible to observe by studying each network node individually.

Hubs: hubs are nodes with a very large number of connections (input and/or output) w.r.t. the average number of connections of the other nodes. In a protein interaction network, where nodes are proteins and two proteins are connected if some kind of direct or indirect interaction between them has been observed in laboratory, hubs are nodes with a lot of different molecular partners, and therefore very likely implied in the majority of the cellular processes. Also, proteins found in many different macromolecular complexes, represented as hubs in the interaction network, could be either a key component of a single molecular complex, or elements shared in many different molecular complexes where they work as switches used to coordinate the activation or repression of different molecular processes. An important consequence of these observations is that any alteration of the hubs of a protein-protein interaction network is predicted to have large effects on the cell biology, in the same way the hacking of a hub server on the internet would cause a wide network malfunction. This assumption, known as "centrality-lethality rule", has been extensively explored in the lab by experimentally knocking-down protein interaction hubs and quantitatively assessing the effects in different models [12]. Going a step further in the reasoning, it has also been hypothesized that mutations

affecting these proteins should be particularly related to the apperance of diseases. Some experimental validation of this prediction has been indeed obtained: Rambaldi and his group provided evidence [18] that virtually all proteins having a degree higher than 80 in the human protein-protein interaction network are target of known cancer-related mutations. Similarly, Ortutay and Vihinen [15], after building an interaction network comprising all human proteins involved in immune response, found that the network hubs included known disease-causing genes as well as 26 new genes related to primary immunodeficiency. In a further example, Chang and colleagues [4], found new gastric cancer candidate markers by looking at hubs in a protein-protein interaction network built from genes differentially expressed in the patient tissues. This is a very good example of a possible role of Systems Biology as an input to the experimental research, and not viceversa.

Cliques: cliques are sets of nodes where each node is connected to all the others. In biological networks cliques proved to be extremely important to identify molecular complexes and/or functional modules. In 2003 Spirin and Mirny wrote a paper [19] describing the presence of densely connected modules in protein-protein interaction networks, i.e., neighborhoods whose internal connectivity was very high compared to the average network connectivity. Thanks to this fact, the authors were able to identify a full set of previously unknown functional modules and molecular machineries. This startling seminal work was very important because it was able to shift the concept of biological network analysis from a single node centrality to a community of nodes. This trend culminated in several complex applications of clique analysis, such as a recent work which nicely illustrates how the mitotic spindle functioning is regulated by a cascade of events which involves cliques (i.e. molecular complexes) instead of single proteins [5].

Graphlets: in easy terms, graphlets are small connected network subgraphs with a predetermined number of nodes. Their biological meaning in networks is relatively new and there are obviously fewer examples. One significant paper discussing the biological meaning of the graphlet degree signature was published by Milenkovic [14]. The results obtained by the authors are somehow superior in generality to other topological features, and uncover the real potential of the network analysis approach in biology. In the paper the authors observe how in a human protein-protein interaction network oncogenes have a very similar graphlet degree signature, which is different from that of genes unrelated to cancer. This observation was so true that it allowed them to use this signature to identify new possible oncogenes. If this finding will be confirmed by other researchers, it will be the proof that the detailed topology around a node in a global protein-protein interaction map is important in determining the function of the corresponding protein at least as its sequence and three-dimensional structure. This result is even more important if we consider the fact that protein-protein interaction networks are only very abstract models of all the interactions which have been observed, without spatial and temporal resolution, and do not correspond to any real physical entity.

B. Analysis of network dynamics

Very interesting biological insights have been recently obtained from the study of the biological network dynamics. For some networks it is possible to define, for each node, a "status", i.e., a condition that can be correlated to a biological condition. For example in Gene Regulatory Networks each node is a gene with an expression level (it's status) corresponding to the gene's transcription rate. The connection between two genes models the causality between them: the expression of gene A directly causes the activation (expression) of gene B. The collection of the status of all the nodes of a network constitutes the *network status*. In biological systems, the need to understand the evolution of the network status (the network dynamics) starting from different initial conditions is becoming evident because, inside a cell, decisions are reached and actions are taken by methods that are exceedingly parallel and extraordinarily integrated.

In biological networks, as well as in many other fields where networks are used for modeling, the key to understand the behavior of these complex systems is to identify and study their *attractors*. An attractor is a state towards a system evolves. Attractors are stable states that "attract" the system dynamics. When a system evolves, the collection of states in which the system transits can either settle to one particular state (point attractor), repeatedly return to a group of states (periodic attractor), or not clearly return to a defined set of states (strange attractor). Chaos and complexity theory are changing the way scientists think about the evolution of complex systems. Recently, the study of attractors, state spaces, and state trajectories has been applied to Gene Regulatory Networks with interesting results which seem to provide a plausible explanation to complex mechanisms, like cell differentiation, and unexpected reactions of tumor cells to standard therapies. For example, in [10] the authors introduce the idea of "Cancer Attractors" that consider tumors as a particular natural state of cell regulation that is normally inaccessible and hidden underneath layers of complex molecular networks. The study of the Gene Regulatory Networks' dynamics could potentially show us how cancer states can be reached, but also how they could be, in some way, avoided. But what is fascinating about this paper is that the idea of Cancer attractors provides a simple formal framework able to explain several complex and unexpected behaviors of cancer cells that traditional biology has not been able to explain yet.

Similarly fascinating works, as the one published by Enver [7], are able to explain how a population of stem cells can diversify in completely different phenotypes (like epithelial cells, nervous cells, cardiac cells, and so on). The idea is that this transition is allowed by the evolution of a regulatory network through several transformations, so that once fully differentiated, it allows the regulatory network to settle into a set of possible attractors that is different for every cell-type, thus allowing the expression of different phenotypes in the cell behavior and morphology.

These pioneering works show that to fully understand the nature of cellular functions, it is necessary to study the behavior of genes in a holistic, rather than in an individual

manner, because the expressions and activities of genes (and consequently of the cell) are not isolated or independent of each other. Moreover, this type of "translation" of well established approaches into a new field of research (Life Sciences) is exactly the type of objective that the engineering community should promote to actively support research in the medical/biological community.

VI. Conclusions

In this paper we discussed some of the methodological challenges that the Life Science world is experiencing following the huge amount of lab data that it is now possible to collect. Understanding the mechanisms that regulate life not only requires more and more advanced lab techniques, but also a shift in the research methodology used to build models and analyize data. The typical bottom-up approach followed by biologists lacks generality as much as the engineers' top-down approach lacks specificity. The optimal solution is probably a new modus operandi able to integrate specific observations made by biologists with functional modeling as seen in engineering.

We are already seeing the first applications of network analysis in human therapy. In particular, although network science is still in its infancy, it is currently shifting from a better understanding of why a given drug works or not, to the identification of new therapeutic interventions. As an example, consider the case of multi-drug therapy, which is a very active field of research and experimental work, due to its high potential in overcoming several obstacle to the effective pharmacological treatment of different conditions. As opposed to the classical "magic bullet" pharmacological paradigm, aiming to the ultra-specific targeting of a single protein, a new kind of approach to the design of a therapy is emerging, which is based on simultaneously targeting several molecular processes. The topological and dynamical analysis of the molecular network underlying a specific disease is the only way to implement such an approach because it allows to look for modulators acting on different network areas so to simultaneously attack different cellular pathways.

In this paper we wanted to discuss some of the methodological issues that, as engineers, we observed when approaching the Life Science world, in order to make readers aware of the great opportunities this research field can offer to the expertese accumulated by researchers in different fields. Understanding the extreme complexity and beauty of life cannot be a task left to biologists, but has to be a multidisciplinary effort, where the best of each research methodology and tool has to be gracefully integrated.

References

[1] Justin Ashworth, Elisabeth J Wurtmann, and Nitin S Baliga. Reverse engineering systems models of regulation: discovery, prediction and mechanisms. *Current Opinion in Biotechnology*, 23(4):598 – 603, 2012.

[2] Albert-Laszlo Barabasi and Zoltan N Oltvai. Network biology: understanding the cell's functional organization. *Nat Rev Genet*, 5(2):101–113, Feb 2004.

[3] Alfredo Benso, Stefano Di Carlo, and Gianfranco Politano. A cDNA Microarray Gene Expression Data Classifier for Clinical Diagnostics Based on Graph Theory. *IEEE/ACM Trans. Comput. Biol. Bioinformatics*, 8(3):577–591, May 2011.

[4] Wenjun Chang, Liye Ma, Liping Lin, Liqiang Gu, Xiaokang Liu, Hui Cai, Yongwei Yu, Xiaojie Tan, Yujia Zhai, Xingxing Xu, Minfeng Zhang, Lingling Wu, Hongwei Zhang, Jianguo Hou, Hongyang Wang, and Guangwen Cao. Identification of novel hub genes associated with liver metastasis of gastric cancer. *Int J Cancer*, 125(12):2844–2853, Dec 2009.

[5] Tzu-Chi Chen, Sheng-An Lee, Chen-Hsiung Chan, Yue-Li Juang, Yi-Ren Hong, Yei-Hsuan Huang, Jin-Mei Lai, Cheng-Yan Kao, and Chi-Ying F Huang. Cliques in mitotic spindle network bring kinetochore-associated complexes to form dependence pathway. *Proteomics*, 9(16):4048–4062, Aug 2009.

[6] Eldad Eilam and Elliot J. Chikofsky. *Reversing: secrets of reverse engineering*. 2007.

[7] Tariq Enver, Martin Pera, Carsten Peterson, and Peter W Andrews. Stem cell states, fates, and the rules of attraction. *Cell Stem Cell*, 4(5):387–397, May 2009.

[8] Aaron P Gabow, Sonia M Leach, William A Baumgartner, Lawrence E Hunter, and Debra S Goldberg. Improving protein function prediction methods with integrated literature data. *BMC Bioinformatics*, 9:198, 2008.

[9] Stefano Gatti, Christian Leo, Simona Gallo, Valentina Sala, Enrico Bucci, Massimo Natale, Daniela Cantarella, Enzo Medico, and Tiziana Crepaldi. Gene expression profiling of hgf/met activation in neonatal mouse heart. *Transgenic Res*, Dec 2012.

[10] Sui Huang, Ingemar Ernberg, and Stuart Kauffman. Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective. *Seminars in cell & developmental biology*, 20(7):869–876, September 2009.

[11] Wolfgang Huber, Vincent J Carey, Li Long, Seth Falcon, and Robert Gentleman. Graphs in molecular biology. *BMC Bioinformatics*, 8 Suppl 6:S8, 2007.

[12] H. Jeong, S. P. Mason, A. L. Barabasi, and Z. N. Oltvai. Lethality and centrality in protein networks. *Nature*, 411(6833):41–42, 05 2001.

[13] Oleksii Kuchaiev and Natasa Przulj. Integrative network alignment reveals large regions of global network similarity in yeast and human. *Bioinformatics*, 27(10):1390–6, May 2011.

[14] Tijana Milenkovic and Natasa Przulj. Uncovering biological network function via graphlet degree signatures. *Cancer Inform*, 6:257–273, 2008.

[15] Csaba Ortutay and Mauno Vihinen. Identification of candidate disease genes by integrating gene ontologies and protein-interaction networks: case study of primary immunodeficiencies. *Nucleic Acids Res*, 37(2):622–628, Feb 2009.

[16] Enrico Pieroni, Sergio de la Fuente van Bentem, Gianmaria Mancosu, Enrico Capobianco, Heribert Hirt, and Alberto de la Fuente. Protein networking: insights into global functional organization of proteomes. *Proteomics*, 8(4):799–816, Feb 2008.

[17] Edi Prifti, Jean-Daniel Zucker, Karine Clément, and Corneliu Henegar. Interactional and functional centrality in transcriptional co-expression networks. *Bioinformatics*, 26(24):3083–9, Dec 2010.

[18] Davide Rambaldi, Federico M Giorgi, Fabrizio Capuani, Andrea Ciliberto, and Francesca D Ciccarelli. Low duplicability and network fragility of cancer genes. *Trends Genet*, 24(9):427–430, Sep 2008.

[19] Victor Spirin and Leonid A Mirny. Protein complexes and functional modules in molecular networks. *Proc Natl Acad Sci U S A*, 100(21):12123–12128, Oct 2003.

[20] B Vogelstein, D Lane, and A J Levine. Surfing the p53 network. *Nature*, 408(6810):307–310, Nov 2000.